

Systematic reviews as a 'lens of evidence': Determinants of benefits and harms of breast cancer screening

Olena Mandrik^{1,2,3}, Nadine Zielonke⁴, Filip Meheus¹, J.L. (Hans) Severens^{2,5}, Neela Guha⁶, Rolando Herrero Acosta¹ and Raul Murillo^{1,7,8}

¹Section of Early Detection and Prevention, International Agency for Research on Cancer, Lyon, France

²Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

³Health Economic and Decision Science (HEDS), School of Health and Related Research (SchARR), The University of Sheffield, Sheffield, United Kingdom

⁴Department of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

⁵Institute for Medical Technology Assessment (iMTA), Erasmus University Rotterdam, Rotterdam, The Netherlands

⁶Section of Evidence Synthesis and Classification, International Agency for Research on Cancer, Lyon, France

⁷Centro Javeriano de Oncología – Hospital Universitario San Ignacio, Bogotá, Colombia

⁸Faculty of Medicine – Pontificia Universidad Javeriana, Bogotá, Colombia

This systematic review, stimulated by inconsistency in secondary evidence, reports the benefits and harms of breast cancer (BC) screening and their determinants according to systematic reviews. A systematic search, which identified 9,976 abstracts, led to the inclusion of 58 reviews. BC mortality reduction with screening mammography was 15–25% in trials and 28–56% in observational studies in all age groups, and the risk of stage III+ cancers was reduced for women older than 49 years. Overdiagnosis due to mammography was 1–60% in trials and 1–12% in studies with a low risk of bias, and cumulative false-positive rates were lower with biennial than annual screening (3–17% vs 0.01–41%). There is no consistency in the reviews' conclusions about the magnitude of BC mortality reduction among women younger than 50 years or older than 69 years, or determinants of benefits and harms of mammography, including the type of mammography (digital vs screen-film), the number of views and the screening interval. Similarly, there was no solid evidence on determinants of benefits and harms or BC mortality reduction with screening by ultrasonography or clinical breast examination (sensitivity ranges, 54–84% and 47–69%, respectively), and strong evidence of unfavourable benefit-to-harm ratio with breast self-examination. The reviews' conclusions were not dependent on the quality of the reviews or publication date. Systematic reviews on mammography screening, mainly from high-income countries, systematically disagree on the interpretation of the benefit-to-harm ratio. Future reviews are unlikely to clarify the discrepancies unless new original studies are published.

Introduction

The traditional evidence-based medicine pyramid places systematic reviews with meta-synthesis on the pinnacle of a hierarchy of evidence. The recently proposed update of the pyramid applies

systematic reviews as a lens through which other types of studies should be appraised, considering synthesised evidence as a tool for stakeholders.¹ But does this lens always provide the same image, and if not, what can affect the conclusions of systematic reviews?

Key words: breast cancer screening, systematic review, benefits, harms, mortality, accuracy, overdiagnosis, false-positive

Abbreviations: AMSTAR: Assessing the Methodological Quality of Systematic Reviews; BCM: breast cancer mortality; BC: breast cancer; BCS: breast cancer screening; BSE: breast self-examination; CBE: clinical breast examination; DCIS: ductal carcinoma *in situ*; FPR: false-positive rates; LMICs: low- and middle-income countries; PPV: positive predictive value; RCTs: randomised controlled trials; RR: relative risk. Additional Supporting Information may be found in the online version of this article.

Grant sponsor: European Commission (OM); **Grant number:** FP7 Marie Curie Actions -People -Co-funding

DOI: 10.1002/ijc.32211

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

History: Received 27 Aug 2018; Accepted 29 Jan 2019; Online 14 Feb 2019

Correspondence to: Olena Mandrik, PhD, Health Economic and Decision Science (HEDS), School of Health and Related Research (SchARR), The University of Sheffield, Regent Court, 30 Regent Street, Sheffield S1 4DA, United Kingdom, Tel.: +44-114-222-4326, Fax: +44-114-222-0749, E-mail: olena.dem@gmail.com; or o.mandrik@sheffield.ac.uk

What's new?

Multiple reviews of the benefits and harms of mammography have been used to inform breast cancer screening guideline development. This process, however, has led to inconsistent screening recommendations. Here, synthesis of results from systematic reviews based on original evidence of determinants of mammography benefits and harms reveals irregularities in data on magnitude of breast cancer reduction obtained with screening mammography. Evidence on determinants of benefits and harms of ultrasonography and clinical breast examination was lacking. Inconsistency in reviews' conclusions was affected by characteristics of the original evidence, indicating that new original studies are needed to clarify discrepancies in screening recommendations.

Many reviews on benefits and harms of breast cancer screening (BCS) have been published over several years. Some of these reviews were used as a basis for developing national or international guidelines, leading to inconsistent recommendations. In a set of systematic reviews, we summarise the data from reviews on four screening approaches – screening mammography, ultrasonography, clinical breast examination (CBE) and breast self-examination (BSE) – or their combinations, among the general population. To our knowledge, no study has previously synthesised the results from systematic reviews on determinants of benefits and harms, participation rate, or cost-effectiveness of BCS approaches or explored the possible differences in the conclusions of systematic reviews on this topic.

In this review, we aim to report:

- (1) Variability in the outcomes of the reviews (mortality reduction, overdiagnosis, false-positive rates (FPR), mortality induced and intermediate outcomes of BCS);
- (2) Variability in the determinants of benefits and harms;
- (3) Review characteristics that explain the variability in the outcomes and derived conclusions.

Methods

The design of this study was reported in the published protocol,² and registered with the International prospective register of systematic reviews (PROSPERO, #CRD42016050764). We systematically searched the PubMed *via* Medline, Scopus, Embase and Cochrane databases in August 2016 and conducted updates and searches for grey literature in February 2017 and again in April 2018 (Appendix 1).

Following the protocol, we excluded reviews not using a systematic (reproducible) literature search. Deviating from the protocol, we included two reviews on which consensus was not reached after two rounds of discussions. For each of the included reviews, we tabulated the outcomes, the score by the Assessing the Methodological Quality of Systematic Reviews (AMSTAR) checklist,³ the limitations of the reviews, and the limitations of the original studies (if their quality was assessed by the reviews and considered in the conclusions). We also narratively summarised the outcomes of the reviews that scored two or higher on the AMSTAR checklist, considering the reviews with lower scores as non-systematic. For the reviews with updates, we synthesised the evidence from the

most recent publication, separately reporting the conclusions of the previous versions.

The uni- and multi-nomial regressions were run in RStudio to assess an impact of factors on the AMSTAR quality score and conclusions of the reviews regarding mammography screening.

Results

We identified 9,976 abstracts through our systematic search and 228 additional reviews through a non-systematic search (Fig. 1). The inter-rater reliability between two reviewers for decisions on full-text inclusion was 85% (Cohen's kappa = 0.63; substantial agreement). The excluded reviews are indicated in Appendix 2.

The 58 included reviews, of which 52 were without updates (Appendix 3), reported data on benefits ($n = 30$), harms ($n = 9$), or both ($n = 19$). Most reviews on benefits and harms of BCS were not limited to a particular geographical region or setting; the others searched for studies comparable to the target countries, such as the UK,^{4–6} the USA,^{7–13} Canada,^{14–17} Australia,¹⁸ the Republic of Korea,¹⁹ or Japan,²⁰ or limited the literature search to a specific region (Asia in one systematic²¹ and Europe in five narrative reviews^{22–26}).

We did not identify systematic reviews reporting the benefits or harms of mammography screening in low- and middle-income countries (LMICs);²⁷ BSE outcomes were reported for China and the Philippines, and CBE outcomes for India. Trials reporting final outcomes of mammography screening, cited in the reviews, were conducted only in, and observational evidence was mainly from, high-income jurisdictions (Appendix 3). A fixed-effects model was used in some of the reviews assessing the clinical outcomes of BCS programmes, including the cluster of Cochrane reviews,^{20,26,28–32} which may signify an assumption of no cross-population differences in the interventions and outcomes.

The structure of the identified outcomes reported in the reviews on benefits and harms of BCS by screening modality is presented in Figure 2.

Screening mammography

Benefits of screening mammography among all age groups. There is consistency on breast cancer mortality (BCM) reduction among meta-analyses (Fig. 3a) and reviews without meta-synthesis (Appendix 4), but no consistency in the interpretation of the size of the effect, the importance of the effect and conclusions on screening with the observed risk or odds ratios

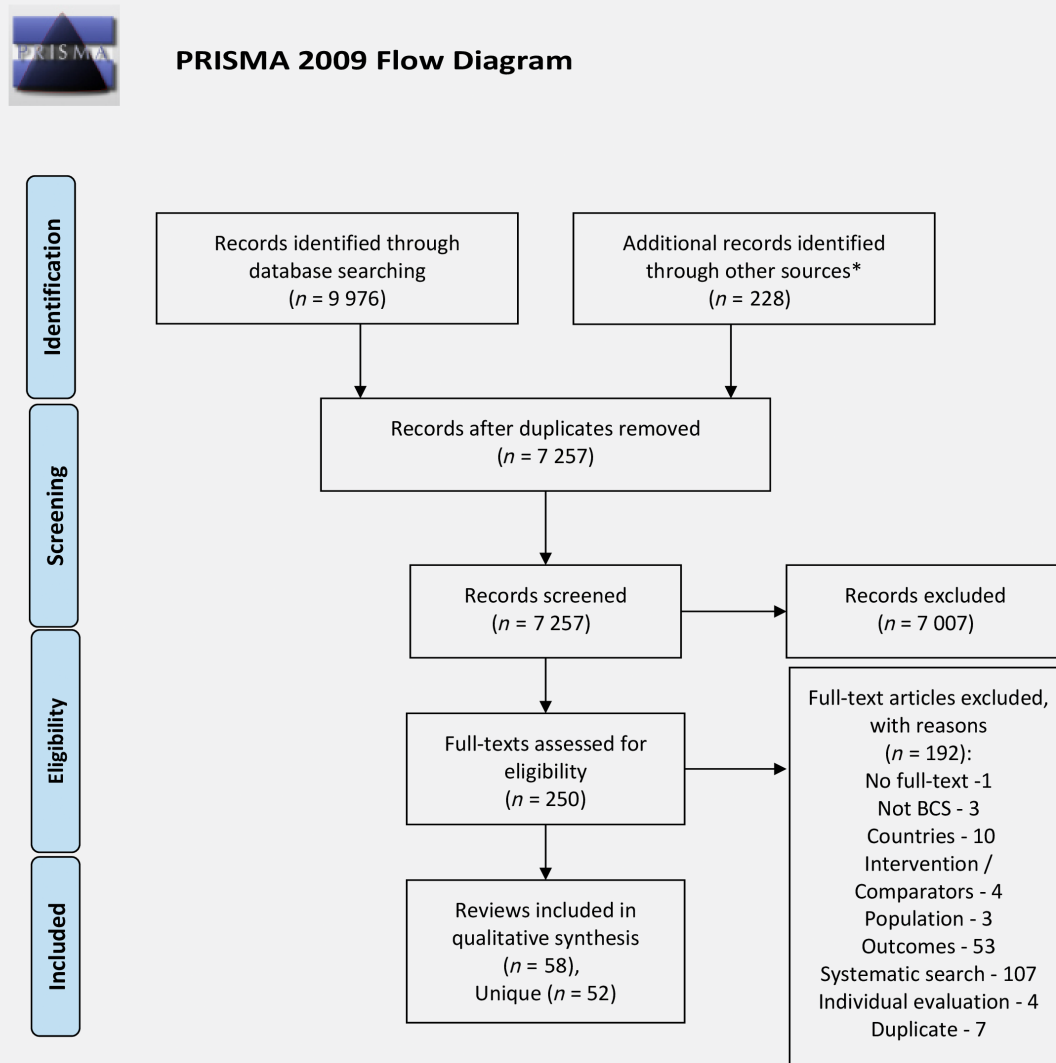


Figure 1. Reproduced with permission from Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA group, *Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement*, PLoS Med, 2009, Vol. 6, page no. e1000097, doi:10.1371/journal.pmed1000097 © PRISMA Statement or PRISMA Explanation and Elaboration. [Color figure can be viewed at wileyonlinelibrary.com]

being justified. The mean size of effect pooled from randomised controlled trials (RCTs) is 15–25%,^{6,9,11,16,19,20,25,30,33–37} from models/estimates is 11–33%,^{33–35} and from observational/population evidence is 28–56%.^{9,25,36} The Cochrane review reported statistically non-significant all-cancer mortality reduction. The all-cause mortality reduction was also statistically non-significant in all the included reviews (Appendix 4).

Overall, the reviews of screening mammography reported high variability of accuracy and intermediate outcomes including sensitivity, size and proportion of small and advanced tumours at diagnosis, proportional interval cancer rate, interval cancer ratio and positive predictive value (PPV). The most

frequently reported outcomes, sensitivity and PPV, had ranges of 51–97% and 2–22%, respectively (Appendix 5).

Although screen-detected tumours may be slow-growing³² and thus lead to overdiagnosis,⁸ tumour size is considered one of the most potent predictors of tumour behaviour in breast cancer (BC).³⁸ The reviews were not fully consistent in concluding that mammography resulted in stage shift or detection of smaller tumours.^{8,12,38–40} We observed that the difference in the conclusions was related to how the target stage shift was defined (stage II+ vs stage III+). No statistically significant relative risk (RR) reduction was observed for shift of stage II+ cancers (Appendix 5). Risk of stage III+ cancers was reduced with

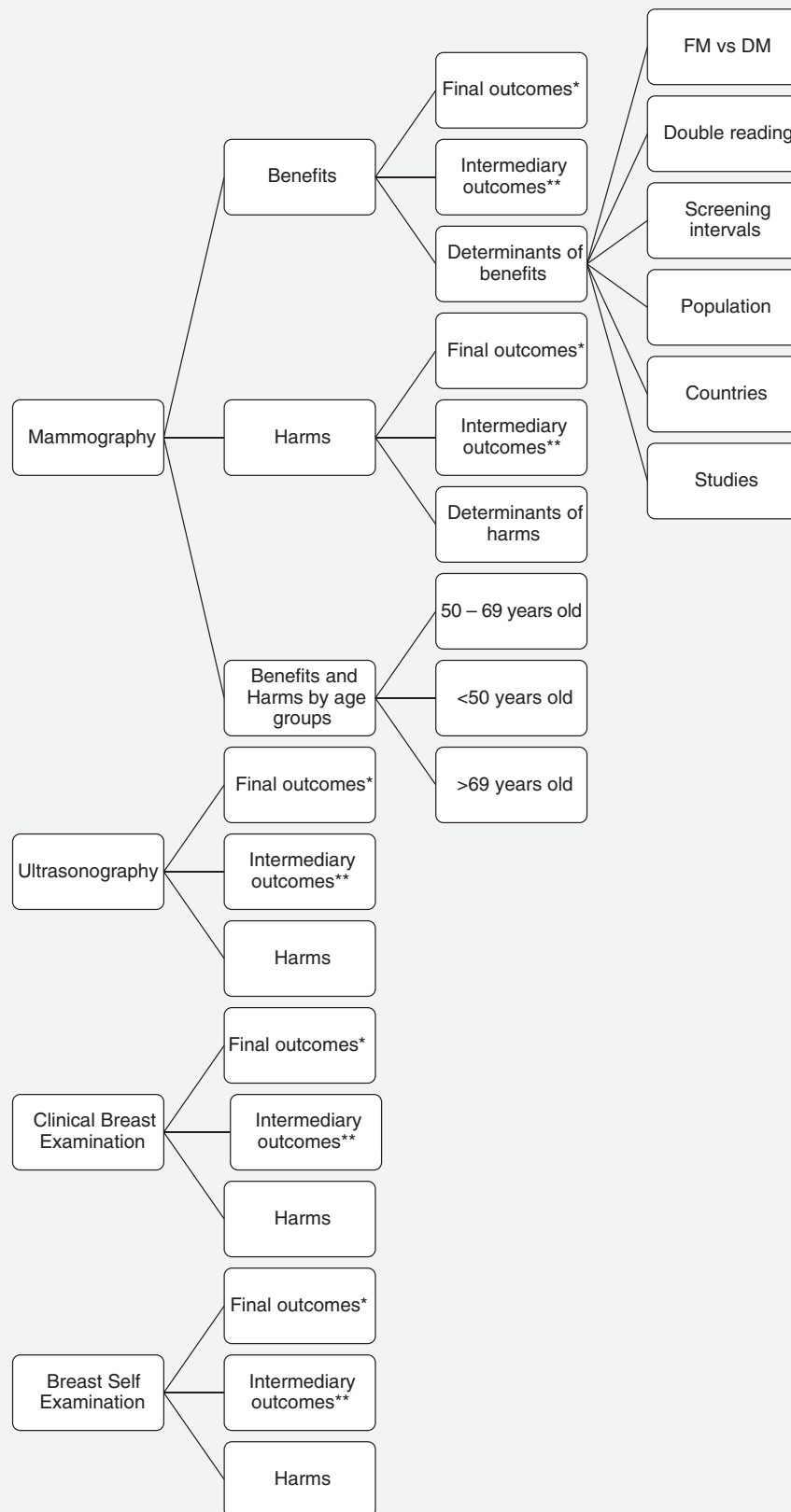


Figure 2. Legend on next page.

mammography screening for women older than 49 years (RR, 0.62; 95% confidence interval, 0.46–0.83) compared to no screening.¹²

Determinants of benefits of screening mammography. There was no consistency in the reviews whether digital mammography has higher or lower accuracy than screen-film mammography (Appendix 6).^{13,15,41–43} The reviews suggested that digital mammography performs better in women younger than 50 years, premenopausal or perimenopausal, with heterogeneously or extremely dense breast tissue.^{15,42} Four reviews concluded on inconsistent evidence on recall rates,^{13,15,41,43} and one on shorter examination times with digital mammography.¹⁵

The included reviews also compared one- vs. two-view mammography, double vs. single reading and screening with different intervals (from 12 to ≥ 36 months). The review by Kerlikowske *et al.* (1995) reported similar BCM reduction with one- and two-view mammography.³⁰ Posso *et al.* (2017)⁴⁴ summarising the evidence from studies where the recall decision was reached by consensus between two readers concluded on similar detection and FPR, while Dinnes *et al.* (2001) suggested that double reading can improve accuracy compared to single reading if a positive decision by any of the readers is sufficient for recall.⁵

Because there were no head-to-head trials comparing effectiveness of BCS by screening intervals, the reviews based their conclusions on indirect comparisons. The conclusions of five reviews were inconsistent about the sufficiency of the evidence on BCM differences with annual vs. biennial or triennial screening.^{9,11,12,16,30} One review found that younger women (<50 years) may benefit more from annual screening, but this evidence was insufficient.⁹

Besides organisational aspects of screening, the reviews also considered breast cancer incidence by age, because higher incidence defined a higher effect of screening, and consistency of effect by country. Humphrey *et al.* (2002) reported that the highest incidence occurred before menopause.⁴⁵

The review of Myers *et al.* (2015) suggested that inconsistency in screening outcomes may be higher in the USA, where there is no single provider for BCS programmes, due to variability between patients, clinicians and insurers.¹¹ Meta-regression analysis of the pooled odds ratios of BCM from case–control studies on BCS did not vary significantly by country.¹⁸

The reviews' conclusions were affected by the characteristics of the original evidence included (trials or observational studies), and by the way the original evidence was analysed and synthesised. There is no observed relationship between initiation dates of RCTs and the reported BCM reduction.¹²

According to Kerlikowske *et al.* (1995), studies initiated before 1980 had lower RR than later studies; the reported confidence intervals of the pooled risk ratios are much wider for later studies than for earlier publications.³⁰ Reviews based on observational evidence report larger BCM reduction than the conclusions based on data from RCTs, with the lowest impact on BCM within the best-randomised trials (Appendix 4).

Harms of screening mammography among all age groups. The main harms reported in the systematic reviews were overdiagnosis, overtreatment related to overdiagnosis, FPR, false-positive biopsies and deaths attributable to radiation induced breast cancer (Appendix 7). The psychological impact of screening is not presented here, because this was not included in the search terms.

Definitions and measurements for overdiagnosis (ranged 0–84%) varied by: type of original evidence, source of cases for the denominator (unscreened, screened detected, entire follow-up, etc.), duration of follow up, accounting for ductal carcinoma *in situ* (DCIS) and other *in situ* lesions, adjustment for breast cancer risk and lead time (Appendix 7). In general, studies using unscreened population in the denominator report higher overdiagnosis and lower rates of overdiagnosis were reported among the pooled values from RCTs and studies with a low risk of bias (Fig. 4):^{6,9–12,16,19,20,28,46} 1–12% (2 of 3 reviews).

Four reviews reported a higher risk of lumpectomies and mastectomies that could be related to a lead-time bias or overdiagnosis.^{12,14,28,47} Screen-detected breast cancers were more frequently treated with radiotherapy,^{12,28,47} but not with chemotherapy or hormone therapy.^{28,47}

Similar to overdiagnosis, FPR and rate of false-positive biopsies varied significantly by screening interval, age of initiation, previous screening experience and source of evidence (Appendix 7). The ranges of non-cumulative FPR were 6.5–8% with annual screening and 1–11% with biennial screening (Appendix 7),^{9,11,47} and of cumulative FPR (after 10 years or lifetime) were 3–63% with annual and 7–60% with biennial screening^{9,12,14,16,19,20,28,37,47} (Fig. 5). Two reviews comparing these screening intervals concluded that FPR is higher with annual screening.^{11,12} The ranges for non-cumulative rate of false-positive biopsies were 2–12% with annual screening^{11,12} and 0.07–9% with biennial screening;^{9,12,14} the cumulative rates (≥ 10 screenings) were 0.01–41% with annual screening^{9,11,47} and 3–17% with biennial screening.^{9,11,47}

In contrast to the other harms, rate of deaths attributable to radiation was not significant in the reviews reporting on the topic.^{12,13,48} Further, the most frequently reported intermediate

Figure 2. Structure of the outcomes of the reviews on benefits and harms of mammography, ultrasonography, clinical breast examination and breast self-examination. *Final outcomes for the benefits of screening: breast cancer mortality, all-cancer mortality, all-cause mortality; final outcomes for harms of screening: overdiagnosis, overtreatment, false-positive diagnosis, and radiation-induced deaths. **Intermediary outcomes for the benefits of screening: sensitivity, size and proportion of small and advanced of tumours at diagnosis, proportional interval cancer rate, interval cancer ratio, positive predictive value; Intermediary outcomes for the harms of screening: specificity, recall rates. Abbreviations: FM, field mammography; DM, digital mammography.

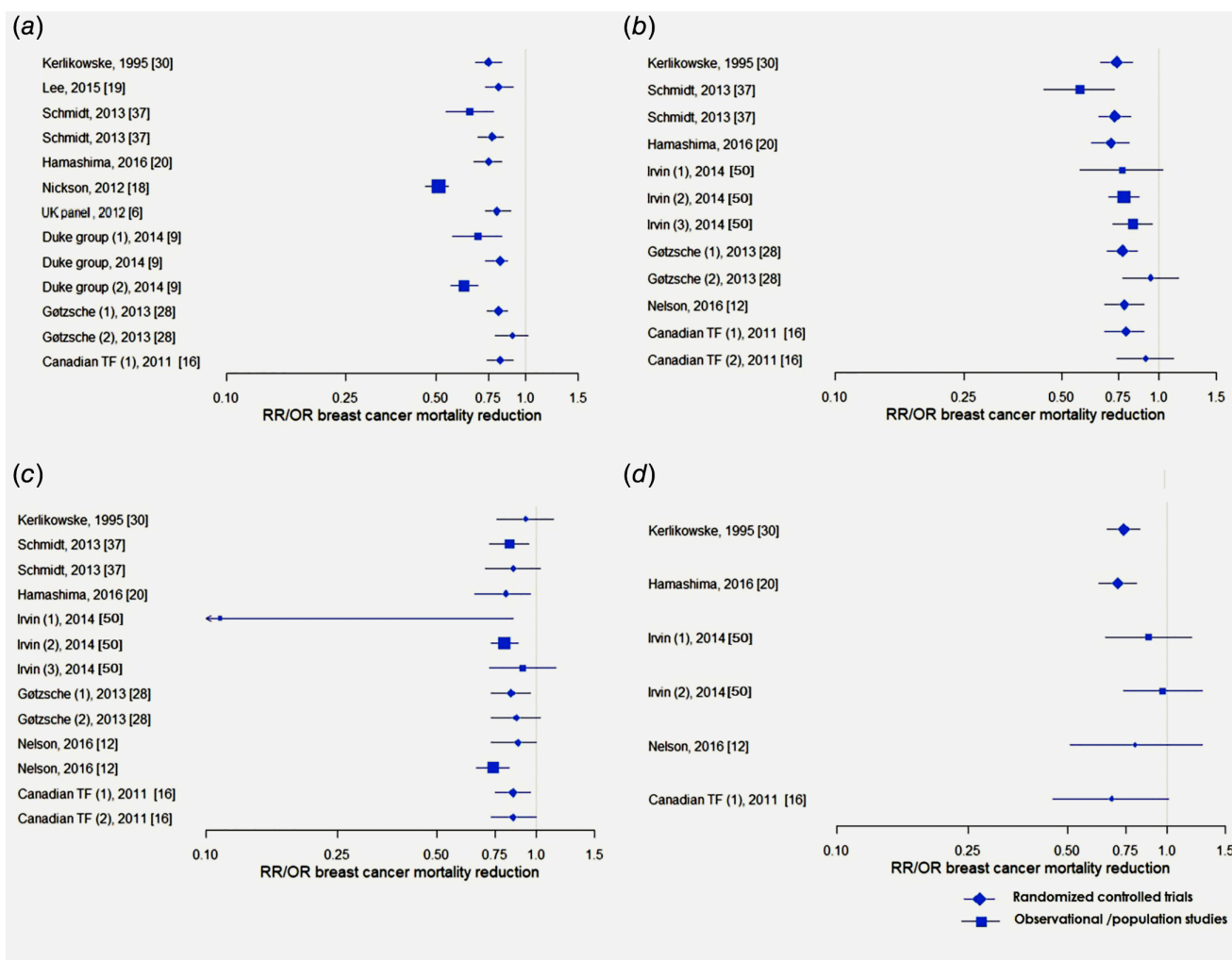


Figure 3. Breast cancer mortality reduction among (a) all age groups, (b) 50–69 year-old, (c) <50 year-old, (d) >69 year-old women. Duke group (2014): (1) Case-control studies, (2) Incidence – based mortality studies; Gotzsche (2013): (1) All randomised trials; (2) Truly randomised trials; Canadian task force (2011): (1) All randomised trials; (2) Truly randomised trials; Irvin (2014): (1) Birth cohort comparison; (2) Geographical comparison; (3) Geographical-Historical Comparisons. [Color figure can be viewed at wileyonlinelibrary.com]

outcomes of harms were specificity (>82% in all the reviews) and recall rate (3–14%) (Appendix 5).

Determinants of harms of screening mammography. Two reviews concluded on limited or no evidence whether overdiagnosis is higher with annual than biennial screening.^{9,11} FPR was considered to be higher with more frequent screenings^{11,12} and with longer duration of screening.¹¹ FPR was also higher for the first screen than for subsequent screens,¹¹ in women with a family history of breast cancer and high breast density, and in women using hormone therapy.¹² The rate of false-positive biopsies per screen decreased with the availability of previous screening results.¹¹ Radiation-related harms increased with higher doses of exposure, younger age at exposure and longer follow-up.⁴⁷

Similar to benefits, harms were not always consistent by country. Several reviews suggested that harms related to BCS

may be higher in the USA,^{7,11,14,20,28} with possible explanations related to different screening and diagnostic guidelines, shorter screening interval, no national provider for screening services and health-care provision through private centres.

Benefits and harms of screening mammography by age groups. Systematic reviews and meta-analyses of the RCTs show a positive effect (22–35%) of mammography screening on BCM reduction among women aged 50–69 years compared to no screening (Fig. 3b, Appendix 4).^{12,13,16,20,28,30,36} All except one systematic review of observational evidence report BCM reduction of 17–49% in this age group.^{12,18,24,26,36,49}

The conclusions and interpretations of the statistical findings of systematic reviews of either RCTs or observational studies reporting BCM reduction among women younger than 50 years^{30,32,45} were and remain inconsistent (Fig. 3c, Appendix 4).^{9,11,12,20,28,37,49} There was no review reporting all-cancer

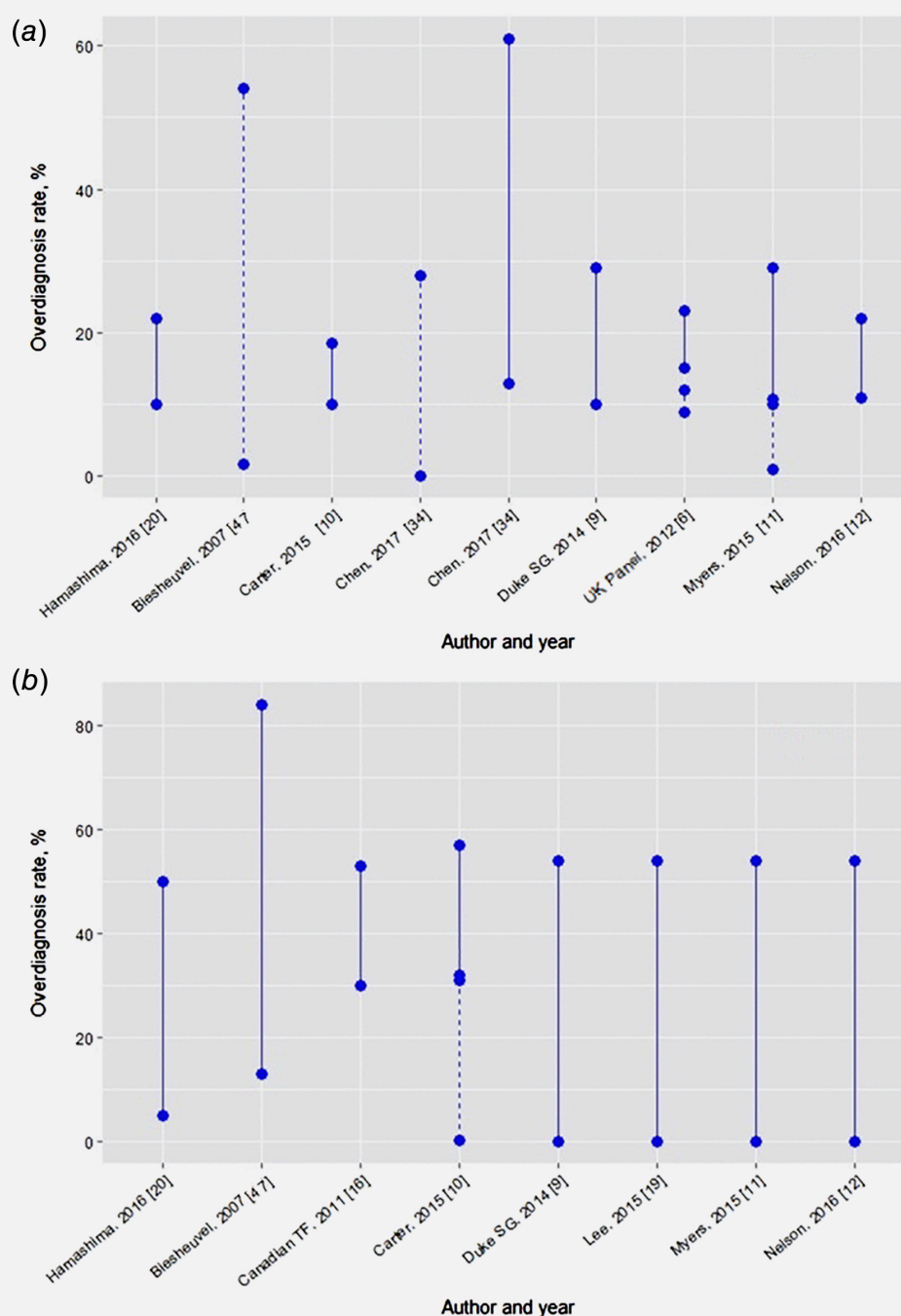


Figure 4. Overdiagnosis rate reported in systematic reviews of (a) randomised controlled trials and (b) observational studies. Dashed line: a—Low risk of bias studies, b—Models. Source of cases in denominators: Hamashima, 2016 [20] - not described; Biesheuvel, 2007 [47]—unscreened; Carter, 2015 - mixed (unscreened - 10%, screen expected - 4-76%, screen detected - 17-31%); Chen, 2017 - unscreened; Duke Synthesis Group, 2014 [9] - not described (unscreened - 29%, screen detected - 19%, entire follow up - 11%); The UK Panel, 2012 [6] - screen detected (16-19%) and entire follow up (10-11%); Myers, 2015 [11] - mixed (screen detected - 19%, entire follow up - 11%); Nelson, 2016 [12] - not described; Canadian Task Force, 2011 [16] - not described; Lee, 2015 [19] - not described. [Color figure can be viewed at wileyonlinelibrary.com]

mortality reduction in this age group, and two meta-analyses concluding on statistically non-significant reduction in all-cause mortality.^{16,37} Seven reviews assumed that mammography

screening has a higher benefit for women older than 50 years and a lower benefit for younger women,^{11-13,20,28,37,47} because of the lower test sensitivity of mammography due to higher breast

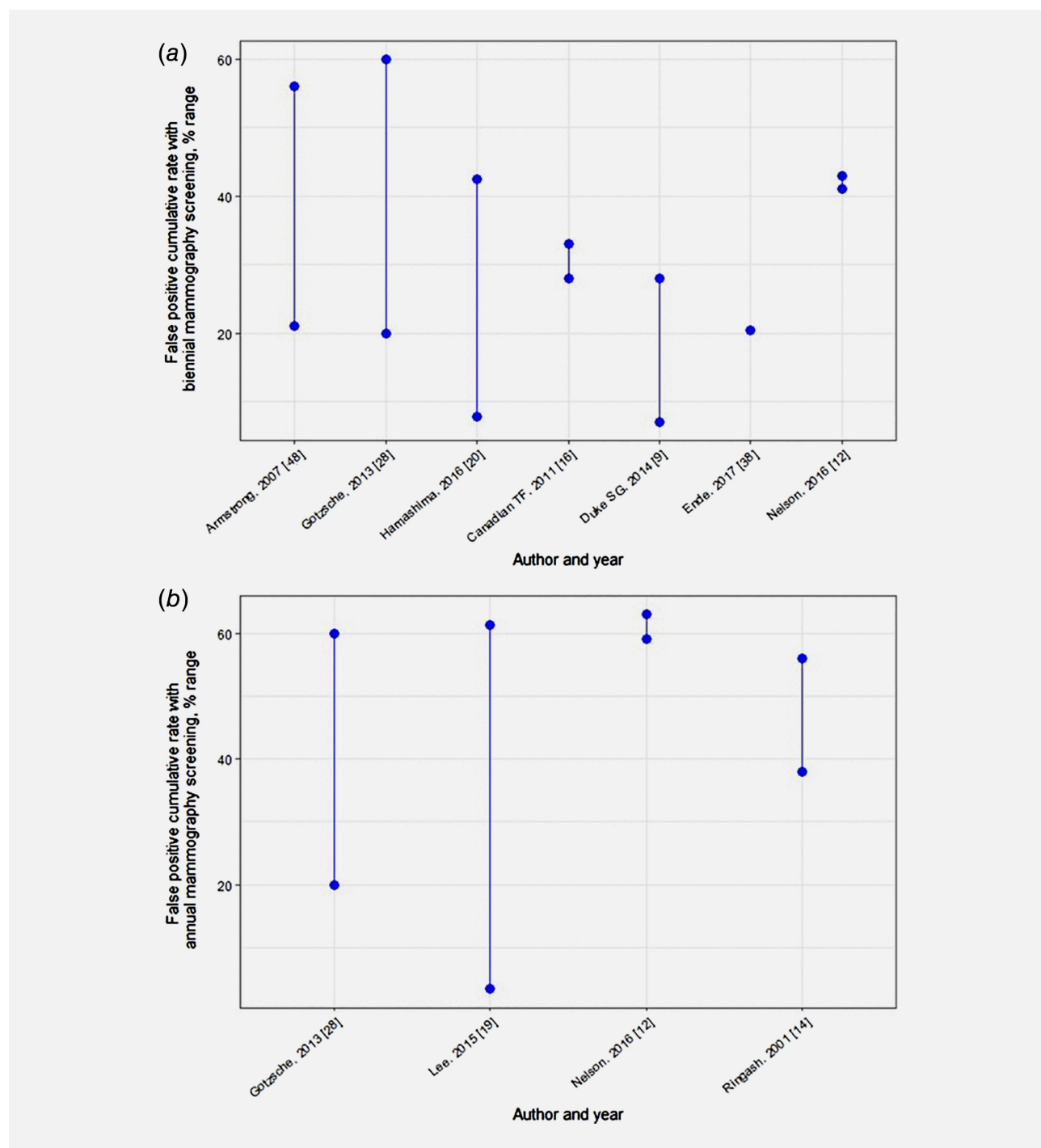


Figure 5. False positive cumulative rates with biennial (a) and annual (b) screening. [Color figure can be viewed at wileyonlinelibrary.com]

density^{12,13,29} and, possibly, faster-growing tumours. Myers *et al.* (2015) suggested that initiating screening at younger ages probably results in greater BCM reduction, but the magnitude of this incremental reduction is uncertain.¹¹ In the high-quality review by Nelson *et al.* (2016),¹² the reduction in risk of advanced stage II+ or stage III+ breast cancers was not statistically significant for women younger than 50 years.

Two included reviews suggest that the rate of overdiagnosis may be larger among women aged 40–49 years,^{12,37} with more than 25% of cases of breast cancer diagnosed among women in their 40s being low-grade DCIS, of which only 14% if left untreated could lead to invasive cancer after several decades.⁴⁷

Although FPR with a single examination was higher for older women,^{11,12} the cumulative FPR was higher among women who

initiated screening early (mainly <50 years).^{12,19,45} The reviews focusing on women younger than 50 years reported cumulative FPR of 20–56%.^{14,37,47} The probability of receiving a certain diagnostic method was age-dependent: women aged 40–49 years experience the highest rate of additional imaging,¹² and therefore may face higher radiation-related harms, whereas their rate of false-positive biopsies is lower than that of older women.¹²

Regarding BCS-induced deaths, several reviews reported limited evidence for women screened annually for 10 years beginning at age 40 years. The estimated number of induced fatal breast cancers is small (8–25 per 100,000 women screened in 3 of 4 reviews)^{12,14,16,45} (Appendix 5), and is higher with earlier initiation of screening.¹²

Similarly to the reviews on younger populations, systematic reviews report inconsistent BCM reduction among women older than 69 years (Fig. 3d, Appendix 4).^{8,11,12,16,20,30,49} A review by Galit *et al.* (2007) concluded on lower BCM among women aged 75–84 years who underwent screening compared to those who did not,⁸ whereas other reviews concluded on no clear benefit for women older than 70 years.^{11,12} Regular mammography has been associated with smaller and earlier-stage tumours among women older than 74 years, which could also be clinically insignificant.⁸ The reviews on BCS benefits and harms among women older than 69 years were based on limited evidence on BCM reduction from RCTs and harms specific to this age group, and did not report all- cancer or all-cause mortality.

Ultrasonography

No high-quality review (out of 6 included) identified studies reporting BCM reduction in BCS among the general population using ultrasonography alone or in combination with mammography (Appendix 8). The reviews targeting Asian populations reported high variability in sensitivity (54–84%), PPV (0.64–6.4%) and FPR (0.9–19.3%) of ultrasonography, with specificity of 96–98% and cancer detection rate of 2–3% per 1,000 screens. The highest-quality reviews^{12,29} concluded that ultrasonography is not justified as a supplementary tool for BCS, because of no solid evidence on its benefits. The reviews did not report transparently which factors can affect the accuracy of ultrasonography.

Clinical breast examination

The 10 included systematic reviews that assessed data on clinical breast examination agreed that the existing data on benefits of CBE are insufficient, because there is no solid evidence on a statistically significant impact of CBE on BCM (Appendix 8).^{9,13,16,20} The range for sensitivity of CBE is 28–36% in the community¹³ and 47–69% in RCTs in all except one review.^{13,19,20,45} The sensitivity of CBE was improved by spending more time on examination and by using a thorough technique.¹³ The specificity of CBE was above 88% in all the reviews.^{13,20,45} compared to no screening, CBE was associated with a higher rate of false-positive biopsies^{13,45} and FPR.^{9,12} No solid evidence was identified on an impact of CBE on life expectancy and overdiagnosis.⁹

Five reviews report no solid evidence on benefits of CBE combined with screening mammography vs. mammography alone.^{9,11,12,20,30} (Appendix 8). The reviews' conclusions varied from "insufficient evidence on effects of CBE" to "no benefits of CBE in terms of mortality reduction"; the review by Lee *et al.* reported an incremental sensitivity of CBE added to mammography of 4–6%, with a decrement in specificity of 2%.¹⁹ Limited data were available on harms of CBE added to mammography, with higher FPR and recall rates reported.^{11,12} Similarly to ultrasonography, the reviews did not report sufficiently on factors affecting the accuracy of CBE, besides an observation of lower sensitivity of screening in real-world vs. trial settings.

Breast self-examination

Six reviews were consistent on no benefit of BSE on BCM (mainly referring to the 3 trials conducted),^{12,13,31,45} all-cause mortality,¹² or number of cancers detected³¹ (Appendix 8). The sensitivity of BSE was 20–41% in a real-world setting vs. 40–89% on silicone models.^{13,17} The specificity of BSE on silicone models was 66–81%.¹⁷ The reviews included reported harms related to FPR, including false-positive biopsies.^{12,13,31}

Quality of the reviews and factors affecting their conclusions

The quality of all of the included reviews varied from 1 to 10 on AMSTAR score (Appendix 9). The reviews were scored the highest on the attributes related to an adequate search approach, description of the included studies and combining the results, and the lowest on reporting conflicts of interest, assessing publication bias, including grey literature and reporting excluded studies (Fig. 6). Multiple regression analysis was used to test if a year of publication, targeting high-income country (vs. none), declaring funding, or including the evidence only from controlled trials significantly predicted AMSTAR

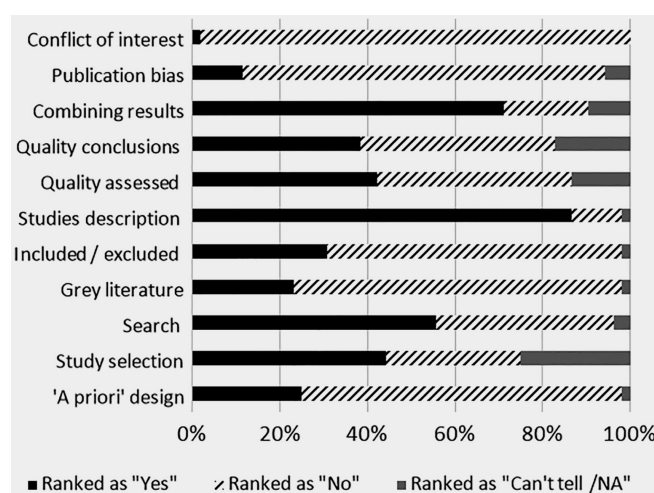


Figure 6. Quality of systematic reviews reporting benefits and/or harms of breast cancer screening.

score of the reviews. The results of the regression indicated that all four factors explained 22% of the variance ($R^2 = 0.22$, $F(6,45) = 2.09$, $p = 0.07$) with funding and target country being not significant factors. The year of publication ($\beta = 0.12$, $p < 0.05$) and type of evidence included ($\beta = -0.82$, $p < 0.05$) explained 16% of variance ($R^2 = 0.16$, $F(2,49) = 4.61$, $p = 0.01$) with model being a better-fit than the univariate analyses.

The results of uni- and multi-variate regressions did not identify significance of such factors as AMSTAR score, date of publication, funding, using qualitative or meta-synthesis, or reporting benefits, harms, or both in the conclusions of the reviews on mammography screening ($p > 0.05$). The conclusions of the reviews reporting similar statistical results were not always identical and may be based on interpretation of statistics, choice of the main outcomes, rigorosity of inclusion criteria and source of evidence. The conclusions of the reviews updated periodically with the new evidence (Appendix 10) did not differ substantially from the previous versions. The publications from one cluster mainly reported similar values for outcomes.

While based on the same RCTs, reviews were inconsistent in the conclusions of trials' biases either in relation to benefits or harms estimation (Appendix 11). The reviews of observational evidence frequently included different studies; the quality of most of them was judged as fair or moderate and the selection bias was the main risk (Appendix 11).

Discussion

Systematic reviews of BCS focus on mammography more than on the other screening approaches, and evaluate benefits of screening more frequently than harms. The available systematic reviews of either benefits or harms of BCS mainly target high-income countries; all RCTs and most of the observational studies on screening mammography were conducted in high-income jurisdictions, on ultrasonography in the USA and Asia, on CBE in North America and Asia, and on BSE in North America, Europe, the Russian Federation and Asia.

The reviews' conclusions on any of the screening approaches were not seen to evolve with time, although some recent updates of the guidelines reported lower importance of mammography screening for younger women compared to earlier versions.^{12,50} We also did not observe a difference in the conclusions of the narrative and systematic reviews. The reviews with high AMSTAR scores and close publication or search date could reach contradictory conclusions on the benefit-to-harm ratio of mammography screening and the justification for its implementation. We found no evidence that variability in the reviews' conclusions was related to objective reasons (search date, rigorosity of inclusion criteria, choice of an outcome, source of evidence). The reviews of more rigorous evidence generally reported both lower benefits and lower harms. We did not see major additive value from the new reviews or updates of the previous reviews on BCS. We conclude that until new high-quality cohort or RCT results are published,

additional reviews on BCS with mammography, ultrasonography, CBE, or BSE would not be of great value.

Summaries of evidence: mammography

The reviews are consistent in reduction in BCM among the general population and women aged 50–69 years, but not all-cancer or all-cause mortality. Both all-cancer and all-cause mortality may serve as the least biased outcomes of the efficacy of screening, avoiding possible mortality misclassifications. However, they may not be sensitive enough to detect the magnitudes in effects. Thus, disease-specific mortality may present the pure effect of the screening programme, while all-cancer and all-cause mortality may be considered in health-care resource allocation and priority setting, enabling comparison of the relative value of screening mammography with other health-care innovations improving survival of the population.

The pure benefits and harms of mammography remain heterogeneous. BCS trials are highly diverse in their protocol designs, adherence and evaluations; combining the outcomes of the RCTs into meta-analyses generates the expectations, but does not predict the outcomes of a specific program (which can either fail or succeed reaching higher effectiveness than meta-synthesised efficacy). Differences between reviews in quality assessment comprise not only identification of bias but also the assignment of overall quality scores, leading to variation in inclusion of RCTs. Subsequently, results of the reviews vary and conclusion were inconsistent. In general, the assessed reviews of RCTs have greater similarity in included studies but larger variability in quality assessment while reviews on observational studies show an opposite trend. If this overview will include only reviews incorporated the quality of studies in their conclusions, the disagreements among the reviews would remain. The impact of screening mammography on stage shift – the most potent intermediate predictor of screening efficacy – was positive for stage III+ breast cancer. BCM increases with progressing tumour stage,⁵¹ and therefore reduction of advanced tumours should improve patients' survival. Tabar *et al.* (2015) calculated that BCM reduction was reaching 28% in the trials achieving 20% or more reduction in advance cancers.⁵² Since BCS programs are long-term planned and costly, detection of advanced cancers should serve as an early indicator of the possible success of the pilot BCS program.

The effectiveness of BCS relates to multiple parameters, including treatment access and efficacy. Regarding access, the health-care settings depicted in RCTs included in systematic reviews may reflect the current situation in LMICs, allowing an approximation of the expected benefits and harms for jurisdictions with limited resources. Furthermore, breast cancer survival also has improved dramatically through the decades due to treatment advances, with age-standardised 5-year survival reaching 85% or higher in 17 high-income countries and 80% or higher in 34 countries worldwide,⁵³ which may diminish the benefits of mammography screening.

If efficacy of late-stage treatments for breast cancer improves more, the clinical benefit of screening may decrease. Concurrently, the accuracy of mammography may also have improved through the years, favouring the benefit-to-harm ratio. Decisions on the rationale for screening should always be a balanced choice of the intervention able to offer the highest benefits with minimum harms, and preferably lower costs.

For women younger than 50 years or older than 69 years, the reviews were not consistent in their conclusions on BCM reduction, with no impact of screening on all-cancer mortality reported. For younger women, most reviews show no impact of mammography screening on early breast cancer detection. The harms may be also higher among younger women (radiation exposure and FPR) and older women (overdiagnosis, because of shorter life expectancy); thus, the evidence collected by included reviews is not consistent on benefit-to-harm ratio for these age groups.

There was no consistency in determinants of higher benefits and lower harms of screening mammography, although double reading may improve sensitivity if the recall decision is based on at least one reader. DCIS is frequently detected and treated during mammography screening. Considering that relative survival with DCIS reached 100% even after 15 years of follow-up,⁵¹ the quality control system should advise on clear and non-aggressive management of screen-detected DCIS. The benefit-to-harm ratio may also be improved with availability of previous screening results. The guidelines on strict quality control and management of non-cancerous lesions could be more important in countries without a national screening provider, like the USA, where harms may be higher than in other countries.

Benefits to harm ratios of mammography screening among women 50 to 59 year old could not remain the same in all jurisdictions. As indicated, effective screening requires organised programs and may vary with disease incidence, population characteristics and structures of financial and health-care systems. Considering the high variability in determinants of benefits and harms of screening, implementing BCS programmes without proper evaluation in these countries is risky, and so the results of the reviews should be extrapolated to LMICs with caution.

For LMICs with high breast cancer incidence and mortality, available early detection programmes, and sufficient capacity, piloting mammography screening among women aged 50–69 years should be combined with evaluation of implementation outcomes before programme scale-up.

Summaries of evidence: ultrasonography, clinical breast examination and breast self-examination

The reviews agree on no solid evidence of mortality reduction with ultrasonography and CBE, and evidence of no effect and higher harms with BSE. Although our review could not summarise evidence from the reviews on reduction in advance breast cancers with CBE, the IARC Handbook on BCS

concluded on sufficient strength of evidence regarding shifts in the stage distribution of tumours detected.⁵⁴ Because mortality reduction with ultrasonography and CBE screening is not confirmed while evidence of potential harms exists, population programmes applying these approaches in countries without access to mammography are questionable. The sensitivity of both methods vary significantly, and real-world implementation may not reach the accuracy reported in trials. The accuracy of these screening approaches is provider-dependent; although CBE is perceived as a low-cost modality, its implementation in communities may entail substantial expenses related to quality assurance, invitations and opportunity costs.

Because of the lack of solid evidence, the benefits and harms of ultrasonography and CBE should be explored further within pilot studies. We consider that appropriate implementation studies on these interventions are necessary even in countries with limited resources, because opportunistic benefits and costs may affect the functioning of the other health programmes.

Research and information gaps

We consider that additional reviews should be discouraged until new original evidence is available. The quality of reviews could be better standardised if the authors were systematically required to apply quality grading instruments to their submitted manuscripts.

More original research on benefits and harms of CBE, ultrasonography and mammography screening among older women is required, which is especially important considering increasing life expectancy. Research targeted at improving the benefit-to-harm ratio of BCS should be encouraged.

The lack of primary and secondary research in LMICs does not enable extrapolation of the evidence to these settings. Because all screening approaches are operator-dependent, high-quality studies are required to gather effectiveness and implementation outcomes of the piloted BCS programmes.

Limitation

Considering the large scope of this systematic review, it is possible that we missed some of the important information despite the comprehensive approach to the evidence search and data extraction. We noted the limitations of using AMSTAR for judging the quality of reviews on cancer screening; some questions on AMSTAR may not be important for reviews of screening studies (such as conflicts of interest of the included studies), low AMSTAR scores may be related to journals' editorial policies on reporting, and high AMSTAR scores may not always mean the absence of biases.

Conclusion

Mammography screening for women aged 50 to 69 years results to decrease in BCM, but not all- cancer and all-cause mortality. It also causes harms, such as overdiagnosis and FPR, which are higher with more frequent screening. The conclusions of the reviews on benefits and harms of mammography were not

consistent for the other age groups. No clear determinants of benefits and harms of mammography screening were identified. The other BCS approaches, such as US, CBE, BSE, cause harms but do not have sufficient evidence on mortality decrease.

Systematic reviews of mammography screening, mainly targeting high-income countries, are discordant in their interpretation of benefits and harms of screening, and their ratio. Their conclusions are not related to their AMSTAR quality score, funding, objectives or the year of publication.

Acknowledgements

The authors are grateful to Taras Vereschak and Kostyantyn Dmitriev, who helped to screen the abstracts, Dr Maribel Almonte, Dr Sabina Rinaldi, Dr Beatrice Lauby-Secretan, Dr Robert Smith and anonymous reviewers who provided valued comments and information regarding the manuscript content, Dr Armando Baena, who advised with graphical presentations, Dr Jin Young Park, Dr Bochen Cao, and Dr Chunqing Lin, who helped with translations, and Dr. Karen Muller who helped to stylistically edit the manuscript.

Disclaimer

The findings and views presented in this manuscript belong to the authors and do not necessarily represent the views of the organisations they are affiliated.

References

- Murad MH, Asi N, Alsawas M, et al. New evidence pyramid. *Evid Based Med* 2016;21:125–7.
- Mandrik O, Ekwunife OI, Zielonke N, et al. What determines the effects and costs of breast cancer screening? A protocol of a systematic review of reviews. *Syst Rev* 2017;6:122.
- Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol* 2009;62:1013–20.
- Petticrew MP, Sowden AJ, Lister-Sharp D, et al. False-negative results in screening programmes: systematic review of impact and implications. *Health Technol Assess (Winchester, England)* 2000; 4:1–120.
- Dinnes J, Moss S, Melia J, et al. Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review. *Breast* 2001;10: 455–63.
- Marmot MG, Altman DG, Cameron DA, et al. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012;380: 1778–86.
- Brewer NT, Salz T, Lillie SE. Systematic review: the long-term effects of false-positive mammograms. *Ann Intern Med* 2007;146:502–10.
- Galit W, Green MS, Lital KB. Routine screening mammography in women older than 74 years: a review of the available data. *Maturitas* 2007;57: 109–19.
- Havrilesky L, Gierisch JM, Moorman P, et al. Systematic review of cancer screening literature for updating American Cancer Society breast cancer screening guidelines. *Duke Evidence Synthesis Group for American Cancer Society* 2014;179.
- Carter JL, Coletti RJ, Harris RP. Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ* 2015;350: g7773.
- Myers ER, Moorman P, Gierisch JM, et al. Benefits and harms of breast cancer screening: a systematic review. *JAMA* 2015;314: 1615–34.
- Nelson HD, Cantor A, Humphrey L, et al. *Screening for Breast Cancer: A Systematic Review to Update the 2009 US Preventive Services Task Force Recommendation*. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville, MD: Agency for Healthcare Research and Quality (US), 2016.
- Elmore JG, Armstrong K, Lehman CD, et al. Screening for breast cancer. *JAMA* 2005;293: 1245–56.
- Ringash J. Preventive health care, 2001 update: screening mammography among women aged 40–49 years at average risk of breast cancer. *Can Med Assoc J* 2001;164:469–76.
- Medical Advisory Secretariat. Cancer screening with digital mammography for women at average risk for breast cancer, magnetic resonance imaging (MRI) for women at high risk: an evidence-based analysis. *Ont Health Technol Assess Ser* 2010;10:1–55.
- Fitzpatrick-Lewis D, Hodgson N, Ciliska D, et al. Breast cancer screening. Canadian Task Force on Preventive Health Care. McMaster University, 2011.
- Baxter N. Preventive health care, 2001 update: should women be routinely taught breast self-examination to screen for breast cancer? *Can Med Assoc J* 2001;164:1837–46.
- Nickson C, Mason KE, English DR, et al. Mammographic screening and breast cancer mortality: a case-control study and meta-analysis. *Cancer Epidemiol Biomark Prev* 2012;21: 1479–88.
- Lee EH, Park P, Kim NS, et al. The Korean guideline for breast cancer screening. *J Korean Med Assoc* 2015;58:408–19.
- Hamashima C, Hamashima CC, Hattori M, et al. The Japanese guidelines for breast cancer screening. *Jpn J Clin Oncol* 2016;46:482–92.
- Huang Y, Pang Y, Wang Q, et al. Evaluation on the accuracy of high-frequency ultrasound being used in the breast cancer screening program in women from Asian countries: a systematic review. *Zhonghua Liu Xing Bing Xue Za Zhi* 2010;31: 1296–9.
- Hofvind S, Ponti A, Patnick J, et al. False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen* 2012;19(Suppl 1):57–66.
- Puliti D, Duffy SW, Miccinesi G, et al. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen* 2012;19(Suppl 1):42–56.
- Broeders M, Moss S, Nystrom L, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. *J Med Screen* 2012;19(Suppl 1): 14–25.
- Moss SM, Nystrom L, Jonsson H, et al. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. *J Med Screen* 2012;19(Suppl 1):26–32.
- Njor S, Nystrom L, Moss S, et al. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen* 2012;19(Suppl 1):33–41.
- World Bank Country and Lending Groups: The World Bank Group, 2018.

Disclosure

The work reported in this paper was undertaken during the tenure of a postdoctoral fellowship of Dr. Olena Mandrik from the International Agency for Research on Cancer, partially supported by the European Commission FP7 Marie Curie Actions, People, Co-funding of regional, national, and international programmes (COFUND).

Definitions

Accuracy—ability of a test to discriminate between the target condition and health, such as sensitivity, specificity and test predictive values;

Ductal carcinoma *in situ*—non-invasive or pre-invasive breast cancer;

False-positive rate—proportion or percentage of screening tests in which a test result improperly indicates presence of breast cancer when in reality it is not present;

Overdiagnosis—the diagnosis of a tumour that would not go on to cause symptoms or death in the woman's lifetime;

Positive Predictive Value—probability that a woman with a positive screening test truly has cancer.

28. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2013;6:CD001877.
29. Gartlehner G, Thaler K, Chapman A, et al. Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk. *Cochrane Database of Syst Rev* 2013;4:CD009632.
30. Kerlikowske K, Grady D, Rubin SM, et al. Efficacy of screening mammography: a meta-analysis. *JAMA* 1995;273:149–54.
31. Kösters J, Gotzsche PC. Regular self-examination or clinical examination for early detection of breast cancer. *Coch Database Syst Rev* 2003;2:CD003373.
32. Olsen O, Gotzsche PC. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2001;58:CD001877.
33. Chen TH, Yen AM, Fann JC, et al. Clarifying the debate on population-based screening for breast cancer with mammography: a systematic review of randomized controlled trials on mammography with Bayesian meta-analysis and causal model. *Medicine* 2017;96:e5684.
34. Koleva-Kolarova RG, Zhan Z, Greuter MJ, et al. Simulation models in population breast cancer screening: a systematic review. *Breast* 2015;24:354–63.
35. Schiller-Fruhworth IC, Jahn B, Arvandi M, et al. Cost-effectiveness models in breast cancer screening in the general population: a systematic review. *Appl Health Econ Health Policy* 2017;15:333–51.
36. Schmidt AF, Rovers MM, Klungel OH, et al. Differences in interaction and subgroup-specific effects were observed between randomized and nonrandomized studies in three empirical examples. *J Clin Epidemiol* 2013;66:599–607.
37. van den Ende C, Oordt-Speets AM, Vrolijk H, et al. Benefits and harms of breast cancer screening with mammography in women aged 40–49 years: a systematic review. *Int J Cancer* 2017;141:1295–306.
38. Nagtegaal ID, Duffy SW. Reduction in rate of node metastases with breast screening: consistency of association with tumor size. *Breast Cancer Res Treat* 2013;137:653–63.
39. Gotzsche PC. Relation between breast cancer mortality and screening effectiveness: systematic review of the mammography trial. *Dan Med Bull* 2011;58:A4246.
40. Autier P, Boniol M, Middleton R, et al. Advanced breast cancer incidence following population-based mammographic screening. *Annals of oncology : official journal of the European Society for Med Oncol* 2011;22:1726–35.
41. Ho C, Hailey D, Warburton R, et al. Digital mammography versus film-screen mammography: technical, clinical and economic assessments. Technology report no 30. *Canadian Coordinating Office for Health Technol Assess* 2002;68.
42. Rothenberg BM, Ziegler KM, Aronson N. Technology evaluation center assessment synopsis: full-field digital mammography. *J Am Coll Radiol* 2006;3:586–8.
43. Iared W, Shigueoka DC, Torloni MR, et al. Comparative evaluation of digital mammography and film mammography: systematic review and meta-analysis. *Sao Paulo Med J* 2011;129:250–60.
44. Posso M, Puig T, Carles M, et al. Effectiveness and cost-effectiveness of double reading in digital mammography screening: a systematic review and meta-analysis. *Eur J Radiol* 2017;96:40–9.
45. Humphrey L, Chan BKS, Detlefsen S, et al. *Screening for Breast Cancer*. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville, MD: Agency for Healthcare Research and Quality (US), 2002.
46. Biesheuvel C, Barratt A, Howard K, et al. Effects of study methods and biases on estimates of invasive breast cancer over detection with mammography screening: a systematic review. *Lancet Oncol* 2007;8:1129–38.
47. Armstrong K, Moye E, Williams S, et al. Screening mammography in women 40 to 49 years of age: a systematic review for the American College of Physicians. *Ann Intern Med* 2007;146:516–26.
48. Erpeldinger S, Fayolle L, Boussageon R, et al. Is there excess mortality in women screened with mammography: a meta-analysis of non-breast cancer mortality. *Trials* 2013;14:368.
49. Irvin VL, Kaplan RM. Screening mammography & breast cancer mortality: meta-analysis of quasi-experimental studies. *Database of Abstracts of Reviews of Effects* 2014;9:e98105. <https://doi.org/10.1371/journal.pone.0098105>.
50. Nelson HD, Tyne K, Naik A, et al. *Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force*. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. Rockville, MD: Agency for Healthcare Research and Quality (US), 2009.
51. Saadatmand S, Bretveld R, Siesling S, et al. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173,797 patients. *BMJ* 2015;351:h4901.
52. Tabar L, Yen AM, Wu WY, et al. Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. *Breast J* 2015;21:13–20.
53. Allemani C, Weir HK, Carreira H, et al. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet* 2015;385:977–1010.
54. IARC Working Group. *Breast cancer screening*. Lyon, France: IARC Working Group on the Evaluation of Cancer-Preventive Interventions, 2014.